

# 3D Computer Vision Models Predict DFT-Level HOMO-LUMO Gap Energies from Force-Field-Optimized Geometries

Michael R. Maser, Sarah E. Reisman  
California Institute of Technology  
Pasadena, CA  
{mmaser, reisman}@caltech.edu

## Abstract

*We investigate 3D deep learning methods for predicting quantum mechanical energies at high-theory-level accuracy from inexpensive, rapidly computed molecular geometries. Using space-filled volumetric representations (voxels), we explore the effects of radial decay from atom centers and rotational data augmentation on learnability. We test several published computer vision models for 3D shape learning, and construct our own architecture based on 3D inception networks with physically meaningful kernels. We provide a framework for further studies and propose a modeling challenge for the computer vision and molecular machine learning communities.*

## 1. Introduction

Obtaining quantum mechanical (QM) properties of molecules with high precision is computationally expensive and scales exponentially with system size [32]. Semi-empirical QM methods offer faster alternatives to density functional theory (DFT) at the expense of accuracy [42]. The combination of low-precision QM modeling with machine learning (ML), so called delta-learning [36, 7, 26, 19], has demonstrated success in obtaining chemically accurate results with minimal additional computational cost. This approach, however, is not amenable to high-throughput experimentation (HTE), e.g., for the in silico screening of a virtual compound library in drug discovery campaigns [32].

The main bottleneck for HTE delta-learning still lies in the optimization of 3D geometry prior to property prediction. On the other hand, structure optimization using simple force field methods [16] can be performed in ultra-high-throughput, though with low accuracy in downstream QM calculations. Therefore, the direct prediction of high-precision QM properties from force-field-optimized geometries represents a valuable goal for molecular ML.

Computer vision (CV) approaches to understanding

molecular geometry have seen little development despite their enormous impact in other 3D shape learning domains [29, 35, 8, 10, 50, 31, 39]. Space-filled volumetric representations match the chemical intuition of molecules as physical objects, though only a handful of reports have utilized voxels [22, 41, 40, 44] or 3D shape descriptors derived from voxels [48] as input features for ML. Even so, studies thus far are limited by use of a single core scaffold for shape differentiation [48], low input resolution at the atomic scale [40], and the prediction of trivially computed properties [22].

We herein report on our initial studies to advance these methods, developing 3D convolutional neural network (CNN) architectures for molecular shape recognition. Our methods are demonstrated by the prediction of QM energy values at near-DFT accuracy using rapidly computed force field geometries [16] rendered as voxel grids. We take advantage of modern graphics processing capabilities for ultra-HT voxelization [44], allowing us to interrogate the effects of atomic radial decay in our input volumes. We find that models based on inception networks [24] provide high accuracy and generalization, and we discuss the chemical relevance of such architectures.

We further find that multi-view augmentation of input data by 10-fold rotation significantly improves accuracy in many cases [29]. We provide a simple workable example for data processing and model construction as an interactive python notebook ([link](#)), and will make the full code base as well as trained models freely available in a follow-up full report. The models are readily usable by researchers in chemistry, and the voxelization and learning functions provide a modeling challenge for the CV community.

## 2. Approach and Methods

### 2.1. Data

The PubChemQC PM6 (PCQCPM6) dataset [32] was recently published and includes over 200 million molecules with geometries optimized at the semi-empirical PM6 the-

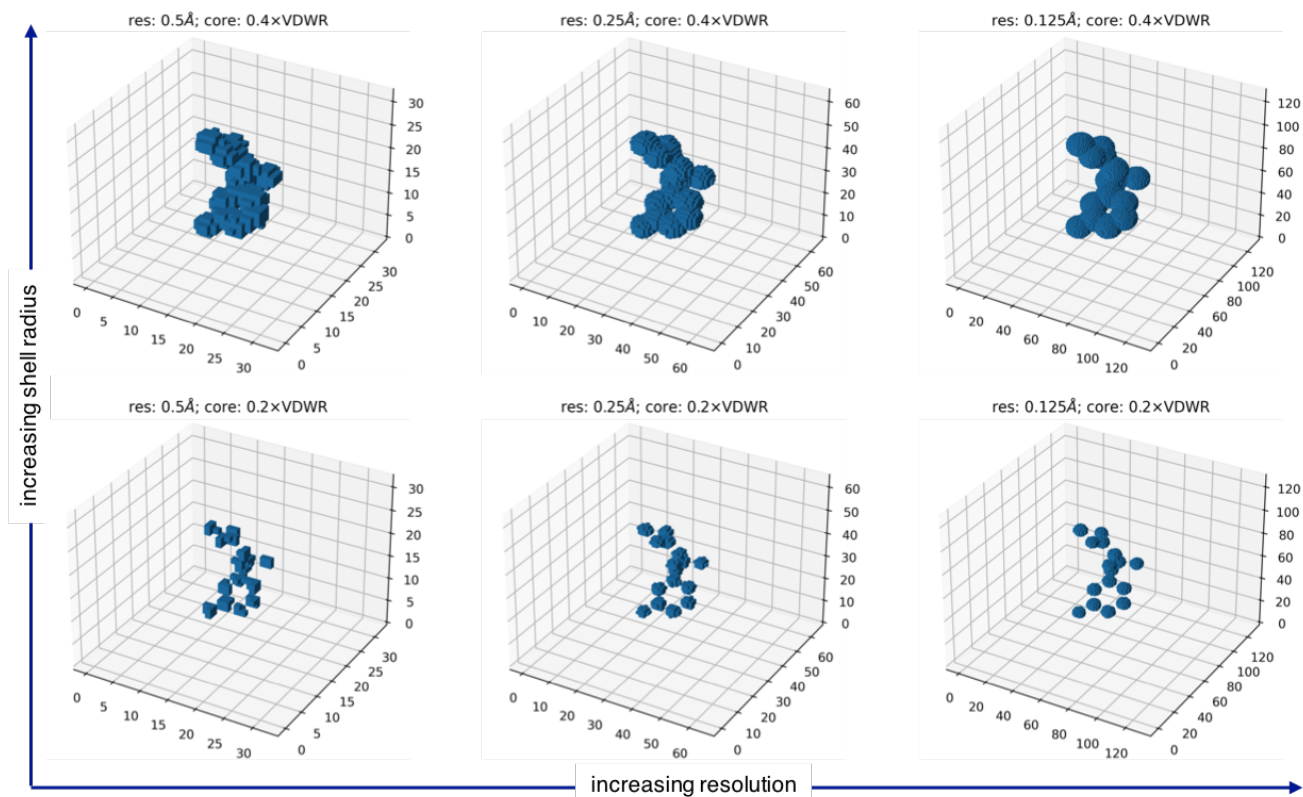


Figure 1. Example renderings of voxelized PCQCPM6 ID: 66098987 (HOMO-LUMO gap 5.76) at various resolutions and decay settings.

ory level [42]. A 2.5 million structure subset was further optimized by high-precision DFT at the B3LYP/6-31G\* theory level. The energy difference between a compound’s highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), so called the HOMO-LUMO gap, is a valuable property often predicted by DFT calculation. The HOMO-LUMO gap is commonly used as an indicator of a substrate’s reactivity, particularly in organic electronics applications [32, 11]. HOMO-LUMO gap energies were published for the entirety of the PCQCPM6 dataset, calculated at the theory level of each compound’s optimization. Given their high accuracy and cost, the B3LYP energy results will serve as our regression targets. These data are rather Gaussian in nature but contain a few significant outliers. Those outside of 5 standard deviations from the mean were removed before modeling (see the original report for more data information [32]).

We prepared space-filled volumes (voxel arrays) for 1.8 million molecules from the B3LYP subset with molecular weights up to 500 g/mol containing the atoms C, H, N, O, P, S, F, and Cl. Geometries were optimized in high throughput using the Merck Molecular Force Field (MMFF) [16] as implemented in the RDKit cheminformatics package [23]. We toggled several settings in the voxelization procedure, in-

cluding the voxel value decay from atom centers as a function of their van der Waals radius (VDWR, see SI).

We additionally tested multiple input resolutions, from  $32^3$  to  $128^3$ , holding the overall grid size constant at  $16\text{\AA}$  in each dimension (Figure 1). We ultimately found higher resolutions offered only minor improvements and for computational speed and data storage settled on  $32^3$  grids, maximum voxel values at 0.3 times the VDWR, stepwise decay in shells of 0.15 times the VDWR, and a total of 4 shells. For the sake of direct comparison between theory-level inputs, we aligned the MMFF structures to the PM6-optimized geometries before modeling (see Section 4).

For architecture assessment, we used a randomly selected subset of 100k molecules. This was divided into training, validation, and test sets with an 80/10/10 split. For the ultimate models, this set was expanded to 1M structures, trained and evaluated with the same 80/10/10 split.

## 2.2. Augmentation

Inspired by performance improvements obtained in the original VoxNet report [29], we experimented with rotational data augmentation. This holds analogy to multi-view 3D object recognition [43], with the extension that our “views” remain in 3D space instead of 2D. To achieve

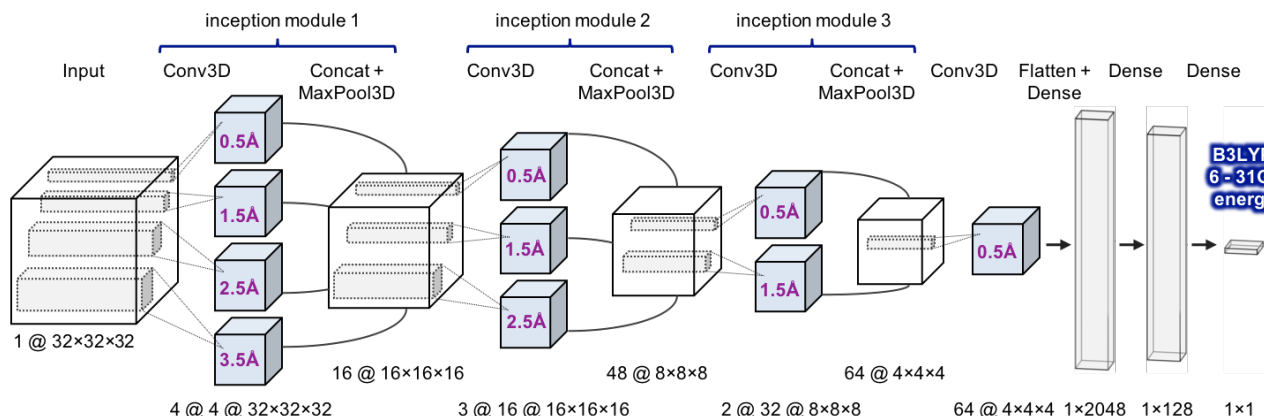


Figure 2. QMCVNet architecture. Uppermost labels represent shorthand layer names. Nethermost labels represent tensor shapes at each layer. Å labels represent the kernel size on one edge used to obtain each intermediate representation in the inception modules.

this, we performed 9 right-angle rotations [4] of each structure (90°, 180°, and 270° about each of the principal axes), yielding 10 total orientations with identical target values when added to the original view. To avoid excessive data storage requirements, we implemented a generator function that performs this augmentation on the fly as examples are passed to the networks (see associated code for details).

### 2.3. Deep learning models

We implemented four published CNN architectures with known success in 3D object recognition [29, 17] and shape-based design [40]. Their composition is described below, and graphical depictions can be found in the SI. Unless specified otherwise, all models were implemented using the TensorFlow deep learning platform [1] and were trained on four NVIDIA® Tesla T4 GPUs for 20 epochs using the Adam optimizer [20] with a learning rate that decayed linearly from  $1e-3$  to  $1e-4$  [29].

**VoxNet.** Reported in 2015 by Maturana and Scherer [29], VoxNet has become a standard baseline architecture for 3D shape learning tasks. It consists of two convolutional layers with kernels, strides, and channels of size (5, 2, 32) and (3, 1, 32), respectively. These are each treated with LeakyReLU activation ( $\alpha = 0.1$ ) and dropout ( $p = 0.2$ ). The outputs of the CNN undergo  $[2 \times 2 \times 2]$  max-pooling and flattening before two ReLU-activated dense layers of output dimensions 128 and 1, the latter providing the predicted HOMO-LUMO gap. In total, VoxNet contains just 0.91M learnable parameters (see SI).

**FusionNet.** An alternative strategy to direct 3D convolutions, FusionNet (Hegde and Zadeh, 2016) [17] takes 2D slices of voxel grids and uses standard image-processing CNN machinery for learning. An interesting consequence of this approach is the anisotropic expansion in the channel dimension during propagation. This translates to stretch-

ing a molecule along one axis as it passes through the network. In the report, “V-CNN I” achieves this by a ReLU-activated CNN layer containing 64 channels with 2D kernels and stride of size (3, 1). 2D max-pooling is followed by two more of the same CNN layers, dropout ( $p = 0.5$ ), and flattening. A dense layer of size 2048 again precedes the output layer. V-CNN I contains 3.4M parameters.

In the same report, “V-CNN II” combines the 2D slice approach with inception modules [24], or multiple kernels of different sizes whose outputs are concatenated and treated together by a nonlinearity in each layer. Specifically, 2D kernels of size 1, 3, and 5 pixels with stride 1 each produce 20 channels (a contraction relative to the size 32 inputs). The three outputs are concatenated to 60 channels before ReLU activation and dropout ( $p = 0.2$ ). Another 2-pronged inception module with kernels of size 1 and 3 combine to give 64 channels, again before ReLU and dropout ( $p = 0.3$ ). A final 32 channel convolution with a  $[3 \times 3]$  kernel, activation, and dropout ( $p = 0.5$ ) is followed by the same dense module as V-CNN I. The largest of all tested architectures, V-CNN II contains over 67M parameters.

**ShapeAutoencoder.** Skalic et al. reported an autoencoder architecture for generating embeddings from 3D molecular shapes. These embeddings are utilized for generative sampling by decoding into SMILES string representations [46]. The encoder architecture is essentially a higher capacity VoxNet, with five feed-forward 3D CNN layers reaching up to 512 channels. Each layer utilizes  $[3 \times 3 \times 3]$  kernels and alternates between strides 1 and 2 for contraction with ReLU activation. After flattening, the embedding is made by a dense layer of size 512; we append a final dense layer to output our scalar predictions. We found the exact implementation of their architecture to give rather unstable loss curves and severe overfitting. We thus added dropout ( $p = 0.2$ ) after the second and fourth CNN layers

Table 1. B3LYP HOMO-LUMO gap prediction performance for all model types on the 100k dataset.

model	theory	test MAE (eV, ↓)	test MAPE (% , ↓)	R <sup>2</sup> (↑)
naive (mean)	–	0.851	16.00	0.000
PM6	–	0.400	7.566	0.783
VoxNet	PM6	0.980	16.84	−0.119
	MMFF	1.256	21.70	−0.719
	MMFF + aug	1.298	22.21	−0.833
V-CNN I	PM6	0.587	11.22	0.504
	MMFF	0.630	11.89	0.426
	MMFF + aug	0.555	10.53	0.553
V-CNN II	PM6	0.577	11.18	0.525
	MMFF	0.610	11.38	0.467
	MMFF + aug	0.595	11.26	0.493
ShapeEncoder-d	PM6	0.539	10.39	0.578
	MMFF	0.554	10.52	0.562
	<b>MMFF + aug</b>	<b>0.502</b>	<b>9.626</b>	<b>0.643</b>
QMCVNet (ours)	PM6	0.543	10.31	0.580
	MMFF	0.549	10.55	0.566
	MMFF + aug	0.530	10.03	0.607

and first dense layer, which led to much smoother fitting and holdout performance (see Section 3 and SI). We refer to the resulting architecture as “**ShapeEncoder-d**” going forward. ShapeEncoder-d contains over 15M parameters.

**QMCVNet (ours).** To derive the benefits of each architecture above, we merged the inception modules of V-CNN II with the deep 3D network of ShapeEncoder-d. For the task of quantum mechanical computer vision, we call our resulting model QMCVNet, depicted in Figure 2. Given that our voxel grids are spaced at 0.5Å, the size 1, 3, and 5 kernels in V-CNN II correspond to 0.5, 1.5, and 2.5Å in each dimension. As typical carbon-carbon and carbon-heteroatom bond lengths are between 1-1.5Å, these kernels convolve over roughly 1-2 bonds, or 2-3 atom centers at most. With the hope of covering entire functional groups (roughly 3-4 atoms), we added a third inception module with kernels of size 1, 3, 5, and 7 to the beginning of the network. After each block, the multi-scale outputs are concatenated and treated with ReLU activation and dropout ( $p = 0.2$ ) before 3D max-pooling of size and stride 2. The output of a final 3D-convolution is flattened and passed through two dense layers of size 2048 and 128 before the HOMO-LUMO gap is predicted. In total, QMCVNet contains 8.7M parameters.

### 3. Results and Analysis

#### 3.1. Architecture screen – 100k examples

To assess their performance, we tested each architecture for predicting HOMO-LUMO gaps at the B3LYP level from single PM6 and MMFF geometries and orientations. Re-

sults from the 100k compound set (see Section 2.1) are included in Table 1. We present model performances as their mean absolute error (MAE) in electron Volts (eV) on the 10k example test set, as this is the metric used in the PCQCPM6 report for comparing the PM6 method against B3LYP itself.

In these experiments, we found that higher capacity models such as ShapeEncoder-d and QMCVNet performed best overall. Smaller and simpler architectures such as VoxNet and V-CNN I tended to underfit the data, yielding mostly mean-centered predictions and unstable loss curves. The best models gave MAEs as low as 0.539 eV for PM6 geometries (ShapeEncoder-d), and 0.549 eV for MMFF (QMCVNet). These MAEs corresponded to mean absolute percentage errors (MAPEs) of around 10.5%, meaning that, on average, predictions were within 10-11% of the true HOMO-LUMO gap values. Correlation coefficients (R<sup>2</sup>) between predictions and ground truths were modest in most cases, topping out around 0.580 (QMCVNet + PM6). Note that the PM6 baseline achieves MAE, MAPE, and R<sup>2</sup> values of 0.400 eV, 7.57%, and 0.783, respectively.

In attempt to increase the performance and generality of our models, we tested the effects of 10-fold rotational augmentation with each model type using the MMFF geometries (see Section 2.2). As in [29], we pass each rotated structure through a given network individually, and take the average of their outputs as the ultimate prediction. Here, we found most models improved over the single-view approach, reducing their MAE by about 5-10%, relatively. ShapeEncoder-d achieved an MAE of 0.502, a MAPE of 9.626, and an R<sup>2</sup> of 0.643.



Table 2. B3LYP HOMO-LUMO gap prediction performance on the 1M dataset.

model	theory	epochs	test MAE (eV, ↓)	test MAPE (% , ↓)	R <sup>2</sup> (↑)
naive (mean)	–	–	0.848	15.86	0.000
PM6	–	–	0.400	7.557	0.782
ShapeEncoder-d	<b>PM6</b>	<b>20</b>	<b>0.418</b>	<b>8.051</b>	<b>0.732</b>
	MMFF	20	0.484	9.196	0.664
	MMFF	100	0.455	8.873	0.687
QMCVNet (ours)	PM6	20	0.479	9.482	0.654
	MMFF	20	0.526	10.38	0.583
	MMFF	100	0.484	9.590	0.645

### 3.2. Best architecture scaleup – 1M examples

With the results above, we tested the best performing architectures—ShapeEncoder-d and QMCVNet—using a dataset of 1M examples, encoded and split as in the previous experiments. Results are included in Table 2, which showed that both models improved substantially with the added data, particularly with PM6 geometries. ShapeEncoder-d achieved a MAE, MAPE, and R<sup>2</sup> of 0.418, 8.051, and 0.732 for PM6 and 0.484, 9.196, and 0.664 for MMFF, respectively. This was particularly exciting as it approaches the accuracy obtained by the true semi-empirical PM6 method, learning from data alone.

If allowed to continue training to 100 epochs, models were able to fit even MMFF geometries down to MAE of 0.455 and MAPE of 8.873 (ShapeEncoder-d). Further, in this experiment, energy evaluations on the 100k compound test set were completed in just 2.5 min, i.e., at a rate of roughly 1.5 ms per prediction. This extrapolates to evaluating the entire 220M PCQCPM6 dataset in under 4 days, compared to the reported 1.5 years for the full calculations [32].

## 4. Discussion and Outlook

There is much room for development based on the work herein. We see several avenues for advancements regarding data, modeling, and applications, each of which are discussed below.

**Data.** In this work, we focused only on single conformers of each example compound, as it significantly reduces data-storage requirements. However, for this and many other 3D learning tasks, evaluating multiple conformers may be critical. Chemical phenomena such as ligand-target binding [44], enantioselective catalysis [45], and structure optimization itself [26] often involve the sampling of many conformers, both (re)active and not. As such, we intend to evaluate the use of conformer ensembles [6] along with our current augmentation techniques to improve on our performance.

Given the Gaussian nature of the B3LYP HL-gap data (see SI), it was challenging to overcome mean-biased predictions, as is the case in many property prediction tasks. Due to the size of our networks and encodings, we elected to retain all data for the sake of effective modeling. However, it is possible that stratified binning and/or over-sampling of data points at the extrema could prove beneficial for overall error reduction.

**Modeling.** Perhaps the most pressing avenue for advancing on our current approach is the direct application of equi- and/or invariant kernel methods. Recent developments have been made in rotationally invariant 3D CNNs [4], which have been applied successfully for medical imaging [3, 5, 47] and texture analysis [25]. Reports thus far have achieved such invariance through the use of rotated convolutional filters [47], spherical harmonics [3], and orientation pooling [4]. These methods are directly applicable to our domain, and represent an attractive approach to molecular shape modeling that could avoid the need for rotational augmentation.

Another interesting approach to improve performance is the use of self- or unsupervised pretraining [27]. In fact, our current models using HOMO-LUMO gap supervision could be used directly as pretrained models for transfer learning with other tasks. We expect, however, that pretraining with an auxiliary task that allows use of the full PCQCPM6 dataset may lead to more transferable models, particularly for tasks in low-data regimes. We intend to explore this avenue first, with regression of the various QM properties contained in the PCQCPM6 examples [32], or even larger descriptor vectors [13] as “self-supervision” tasks [27].

Lastly, given the results observed herein, much simpler architecture and hyperparameter adjustments may give some performance boosts. In particular, our QMCVNet is just  $\frac{1}{2}$  and  $\frac{1}{8}$  the size of ShapeEncoder-d and V-CNN II, respectively, but still performs nearly as well or better in most cases. As such, simply expanding the inception modules or dense layers may prove fruitful, as could neural architecture search [12]. We leave these studies for future work.

**Applications.** Ultimately, it is our goal to develop 3D CNNs for the understanding of organic reactivity. At the fundamental level, reactions proceed through 3D interactions, and we therefore expect that explicitly modeling the 3D shape of reaction components could lead to better performance for reactivity tasks. These include but are not limited to reaction yield [37, 51], selectivity [49, 15], and condition prediction [28], and even retrosynthesis planning [38].

It should be noted that simpler ML methods using, e.g., fingerprints or descriptors with tree models or feed-forward NNs [33, 2], or graph convolutional networks [18, 21, 14], may be better suited for each of these tasks, including that presented here. This is particularly true for small datasets, where volumetric encodings may be unreasonably large, and for problems where electronic structure may outweigh steric shape. However, with advancements in high-throughput experimentation [9, 34, 30], pretraining methods [27], and graphics processing and data storage [44], we expect that 3D learning may become achievable for these inherently 3D tasks, and hope that our current work will inform future studies.

## 5. Conclusion

Herein, we presented results from 3D deep learning experiments with voxel representations for predicting high-theory-level quantum mechanical properties from rapidly computed force-field geometries. We investigated several state-of-the-art 3D CNN models, and developed our own architecture—QMCVNet—combining concepts from inception networks [24] and 3D shape recognition [17]. We found that high capacity models tended to perform best, giving accuracies near semi-empirical quantum methods and with dramatic increase in speed. We presented several opportunities for further improvement on our results, which we are currently investigating. It is our hope that our developed methods and the availability of the PubChemQC PM6 dataset [32] will lead to community engagement to advance research at the intersection of computer vision and chemistry.

## Acknowledgements

Fellowship support was provided by the NSF (M.R.M., Grant No. DGE-1144469). S.E.R. is a Heritage Medical Research Investigator. Financial support from the Research Corporation Cottrell Scholars Program is acknowledged.

## References

[1] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur,

Josh Levenberg, Rajat Monga, Sherry Moore, Derek G Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: A system for large-scale machine learning. page 21, Nov. 2016. 3

[2] Derek T. Ahneman, Jesús G. Estrada, Shishi Lin, Spencer D. Dreher, and Abigail G. Doyle. Predicting Reaction Performance in C–N Cross-Coupling Using Machine Learning. *Science*, 360(6385):186–190, Apr. 2018. 6

[3] Vincent Andrearczyk, Julien Fageot, Valentin Oreiller, Xavier Montet, and Adrien Depeursinge. Exploring local rotation invariance in 3D CNNs with steerable filters. In *Proceedings of Machine Learning Research*, volume 102, pages 15–26, 2019. 5

[4] Vincent Andrearczyk, Julien Fageot, Valentin Oreiller, Xavier Montet, and Adrien Depeursinge. Local Rotation Invariance in 3D CNNs. *arXiv:2003.08890 [cs, eess]*, Mar. 2020. Preprint. 3, 5

[5] Vincent Andrearczyk, Valentin Oreiller, Julien Fageot, Xavier Montet, and Adrien Depeursinge. Solid Spherical Energy (SSE) CNNs for Efficient 3D Medical Image Analysis. page 9, Technological University Dublin, Dublin, Ireland, 2019. 5

[6] Simon Axelrod and Rafael Gomez-Bombarelli. Molecular machine learning with conformer ensembles. *arXiv:2012.08452 [physics]*, Feb. 2021. Preprint. 5

[7] Mihail Bogojeski, Leslie Vogt-Maranto, Mark E. Tuckerman, Klaus-Robert Müller, and Kieron Burke. Quantum chemical accuracy from density functional approximations via machine learning. *Nature Communications*, 11(1):5223, Dec. 2020. 1

[8] Andrew Brock, Theodore Lim, J. M. Ritchie, and Nick Weston. Generative and Discriminative Voxel Modeling with Convolutional Neural Networks. *arXiv:1608.04236 [cs, stat]*, Aug. 2016. arXiv: 1608.04236. 1

[9] Alexander Buitrago Santanilla, Erik L. Regalado, Tony Pereira, Michael Shevlin, Kevin Bateman, Louis-Charles Campeau, Jonathan Schneeweis, Simon Berritt, Zhi-Cai Shi, Philippe Nantermet, Yong Liu, Roy Helmy, Christopher J. Welch, Petr Vachal, Ian W. Davies, Tim Cernak, and Spencer D. Dreher. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science*, 347(6217):49–53, Jan. 2015. 6

[10] Ali Caglayan and Ahmet Burak Can. Volumetric Object Recognition Using 3-D CNNs on Depth Data. *IEEE Access*, 6:20058–20066, 2018. 1

[11] Weiran Cao and Jiangeng Xue. Recent progress in organic photovoltaics: device architecture and optical design. *Energy Environ. Sci.*, 7(7):2123, 2014. 2

[12] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. Neural Architecture Search: A Survey. *arXiv:1808.05377 [cs, stat]*, Apr. 2019. Preprint. 5

[13] Benedek Fabian, Thomas Edlich, Hélène Gaspar, Marwin Segler, Joshua Meyers, Marco Fiscato, and Mohamed Ahmed. Molecular representation learning with language models and domain-relevant auxiliary tasks. *arXiv:2011.13230 [cs]*, Nov. 2020. arXiv: 2011.13230. 5

- [14] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural Message Passing for Quantum Chemistry. *arXiv:1704.01212 [cs]*, Preprint, June 2017. **6**
- [15] Yanfei Guan, Connor W. Coley, Haoyang Wu, Duminda Ranasinghe, Esther Heid, Thomas J. Struble, Lagnajit Patanaik, William H. Green, and Klavs F. Jensen. Regio-selectivity prediction with a machine-learned reaction representation and on-the-fly quantum mechanical descriptors. *Chem. Sci.*, 12(6):2198–2208, 2021. **6**
- [16] Thomas A. Halgren. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *Journal of Computational Chemistry*, 17(5-6):490–519, 1996. **1, 2**
- [17] Vishakh Hegde and Reza Zadeh. FusionNet: 3D Object Classification Using Multiple Data Representations. *arXiv:1607.05695 [cs.CV]*, Preprint:9, 2016. **3, 6**
- [18] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular Graph Convolutions: Moving Beyond Fingerprints. *J. Comput.-Aided Mol. Des.*, 30(8):595–608, Aug. 2016. **6**
- [19] John A. Keith, Valentin Vassilev-Galindo, Bingqing Cheng, Stefan Chmiela, Michael Gastegger, Klaus-Robert Müller, and Alexandre Tkatchenko. Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems. *Chem. Rev.*, page acs.chemrev.1c00107, July 2021. **1**
- [20] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, Preprint, Jan. 2017. **3**
- [21] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv:1609.02907 [cs, stat]*, Preprint, Feb. 2017. **6**
- [22] Denis Kuzminykh, Daniil Polykovskiy, Artur Kadurin, Alexander Zhebrak, Ivan Baskov, Sergey Nikolenko, Rim Shayakhmetov, and Alex Zhavoronkov. 3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks. *Mol. Pharmaceutics*, 15(10):4378–4385, Oct. 2018. **1**
- [23] Gregory A. Landrum. RDKit: Open-Source Cheminformatics Software. (accessed Nov 20, 2018). **2**
- [24] Min Lin, Qiang Chen, and Shuicheng Yan. Network In Network. *arXiv:1312.4400 [cs]*, Mar. 2014. *arXiv: 1312.4400*. **1, 3, 6**
- [25] Li Liu, Jie Chen, Paul Fieguth, Guoying Zhao, Rama Chellappa, and Matti Pietikäinen. From BoW to CNN: Two Decades of Texture Representation for Texture Classification. *Int J Comput Vis*, 127(1):74–109, Jan. 2019. **5**
- [26] Jianing Lu, Song Xia, Jieyu Lu, and Yingkai Zhang. Dataset Construction to Explore Chemical Space with 3D Geometry and Deep Learning. *J. Chem. Inf. Model.*, page acs.jcim.1c00007, Mar. 2021. **1, 5**
- [27] Huanru Henry Mao. A Survey on Self-supervised Pre-training for Sequential Transfer Learning in Neural Networks. *arXiv:2007.00800 [cs, stat]*, July 2020. Preprint. **5, 6**
- [28] Michael R. Maser, Alexander Y. Cui, Serim Ryou, Travis J. DeLano, Yisong Yue, and Sarah E. Reisman. Multilabel Classification Models for the Prediction of Cross-Coupling Reaction Conditions. *J. Chem. Inf. Model.*, 61(1):156–166, Jan. 2021. **6**
- [29] Daniel Maturana and Sebastian Scherer. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 922–928, Hamburg, Germany, Sept. 2015. IEEE. **1, 2, 3, 4**
- [30] Steven M. Mennen, Carolina Alhambra, C. Liana Allen, Mario Barberis, Simon Berritt, Thomas A. Brandt, Andrew D. Campbell, Jesús Castañón, Alan H. Cherney, Melodie Christensen, David B. Damon, J. Eugenio de Diego, Susana García-Cerrada, Pablo García-Losada, Rubén Haro, Jacob Janey, David C. Leitch, Ling Li, Fangfang Liu, Paul C. Lobben, David W. C. MacMillan, Javier Magano, Emma McInturff, Sebastien Monfette, Ronald J. Post, Danielle Schultz, Barbara J. Sitter, Jason M. Stevens, Iulia I. Strambeanu, Jack Twilton, Ke Wang, and Matthew A. Zajac. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.*, 23(6):1213–1242, June 2019. **6**
- [31] A. A. M. Muzahid, Wanggen Wan, Ferdous Sohel, Naimat Ullah Khan, Ofelia Delfina Cervantes Villagomez, and Hidayat Ullah. 3D Object Classification Using a Volumetric Deep Neural Network: An Efficient Octree Guided Auxiliary Learning Approach. *IEEE Access*, 8:23802–23816, 2020. **1**
- [32] Maho Nakata, Tomomi Shimazaki, Masatomo Hashimoto, and Toshiyuki Maeda. PubChemQC PM6: Data Sets of 221 Million Molecules with Optimized Molecular Geometries and Electronic Properties. *J. Chem. Inf. Model.*, 60(12):5891–5899, Dec. 2020. **1, 2, 5, 6**
- [33] Matthew K. Nielsen, Derek T. Ahneman, Orestes Riera, and Abigail G. Doyle. Deoxyfluorination with Sulfonyl Fluorides: Navigating Reaction Space with Machine Learning. *J. Am. Chem. Soc.*, 140(15):5004–5008, Apr. 2018. **6**
- [34] Damith Perera, Joseph W. Tucker, Shalini Brahmabhatt, Christopher J. Helal, Ashley Chong, William Farrell, Paul Richardson, and Neal W. Sach. A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science*, 359(6374):429–434, Jan. 2018. **6**
- [35] Charles R. Qi, Hao Su, Matthias Niessner, Angela Dai, Mengyuan Yan, and Leonidas J. Guibas. Volumetric and Multi-View CNNs for Object Classification on 3D Data. *arXiv:1604.03265 [cs]*, Apr. 2016. *arXiv: 1604.03265*. **1**
- [36] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Big Data Meets Quantum Chemistry Approximations: The -Machine Learning Approach. *J. Chem. Theory Comput.*, 11(5):2087–2096, May 2015. **1**
- [37] Philippe Schwaller, Alain C Vaucher, Teodoro Laino, and Jean-Louis Reymond. Prediction of chemical reaction yields using deep learning. *Mach. Learn.: Sci. Technol.*, 2(1):015016, Mar. 2021. **6**
- [38] Marwin H. S. Segler, Mike Preuss, and Mark P. Waller. Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI. *Nature*, 555(7698):604–610, Mar. 2018. **6**

- [39] Satya P. Singh, Lipo Wang, Sukrit Gupta, Haveesh Goli, Parasuraman Padmanabhan, and Balázs Gulyás. 3D Deep Learning on Medical Images: A Review. *Sensors*, 20(18):5097, Sept. 2020. [1](#)
- [40] Miha Skalic, José Jiménez, Davide Sabbadin, and Gianni De Fabritiis. Shape-Based Generative Modeling for de Novo Drug Design. *Journal of Chemical Information and Modeling*, Feb. 2019. [1](#), [3](#)
- [41] Miha Skalic, Alejandro Varela-Rial, José Jiménez, Gerard Martínez-Rosell, and Gianni De Fabritiis. LigVoxel: inpainting binding pockets using 3D-convolutional neural networks. *Bioinformatics*, 35(2):243–250, Jan. 2019. [1](#)
- [42] James J. P. Stewart. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J Mol Model*, 13(12):1173–1213, Oct. 2007. [1](#), [2](#)
- [43] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view Convolutional Neural Networks for 3D Shape Recognition. *arXiv:1505.00880 [cs]*, Sept. 2015. arXiv: 1505.00880. [2](#)
- [44] Jocelyn Sunseri and David R. Koes. libmolgrid: Graphics Processing Unit Accelerated Molecular Gridding for Deep Learning Applications. *Journal of Chemical Information and Modeling*, 60(3):1079–1084, Mar. 2020. [1](#), [5](#), [6](#)
- [45] Konstantinos D. Vogiatzis, Mikhail V. Polynski, Justin K. Kirkland, Jacob Townsend, Ali Hashemi, Chong Liu, and Evgeny A. Pidko. Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities. *Chem. Rev.*, 119(4):2453–2523, Feb. 2019. [5](#)
- [46] David Weininger. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.*, 28(1):31–36, Feb. 1988. [3](#)
- [47] Marysia Winkels and Taco S. Cohen. Pulmonary nodule detection in CT scans with equivariant CNNs. *Medical Image Analysis*, 55:15–26, July 2019. [5](#)
- [48] Andrew F. Zahrt, Jeremy J. Henle, Brennan T. Rose, Yang Wang, William T. Darrow, and Scott E. Denmark. Prediction of Higher-Selectivity Catalysts by Computer-Driven Workflow and Machine Learning. *Science*, 363(6424):eaau5631, Jan. 2019. [1](#)
- [49] Shibin Zhao, Tobias Gensch, Benjamin Murray, Zachary L. Niemeyer, Matthew S. Sigman, and Mark R. Biscoe. Enantiodivergent Pd-catalyzed C–C bond formation enabled through ligand parameterization. *Science*, 362(6415):670–674, Nov. 2018. [6](#)
- [50] Shuaifeng Zhi, Yongxiang Liu, Xiang Li, and Yulan Guo. Toward real-time 3D object recognition: A lightweight volumetric CNN framework using multitask learning. *Computers & Graphics*, 71:199–207, Apr. 2018. [1](#)
- [51] Andrzej M. Żurański, Jesus I. Martinez Alvarado, Benjamin J. Shields, and Abigail G. Doyle. Predicting Reaction Yields via Supervised Learning. *Acc. Chem. Res.*, 54(8):1856–1865, Apr. 2021. [6](#)